

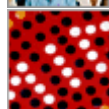
Summary

Main Features

CONTENTS



PREFACE



INTRODUCTION



CENSUS DATA ENHANCEMENT PROJECT - 2006



CENSUS DATA ENHANCEMENT PROJECT - 2011



CONFIDENTIALITY AND PRIVACY



APPENDIX 1 - CENSUS DATA ENHANCEMENT PROJECT 2006

About this Release

Commencing with the 2006 Census, the ABS began a project to enhance the value of Census data by bringing it together with other datasets to leverage more information from the combination of individual datasets than is available from the datasets separately.

This paper provides an update with respect to the project for the 2011 Census of Population and Housing.

Expanded Contents

CONTENTS

Preface

Introduction

Census Data Enhancement Project - 2006

Census Data Enhancement Project - 2011

1 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, for quality studies

2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs

2.1 Indigenous Mortality Project

2.2 Enhancing Australia's Cancer Statistics Project

3 Wave 2 of a 5% Statistical Longitudinal Census Dataset

4 Bringing together the SLCD with other datasets without using name and address for statistical and research purposes

5 Bringing together 2011 Census data with other datasets without using name and address after Census processing

Confidentiality and Privacy

Legislative Protection

Destruction of census forms and name and address information

Access to ABS information

Data security

Appendix

Preface

Contents >> Preface



PREFACE

In 2005, the ABS embarked on a project to enhance the value of Census data by bringing it together with other datasets, both ABS and non-ABS, to leverage more information from the combination of individual datasets than is available from the individual datasets separately.

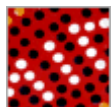
The Census Data Enhancement (CDE) project improves and expands the range of official statistics available to Australian society, and improves the evidence base to support good government policy making, program management and service delivery.

This paper provides an update on the CDE project for the 2011 Census of Population and Housing.

Brian Pink
Australian Statistician

Introduction

[Contents >> Introduction](#)



INTRODUCTION

The Census Data Enhancement (CDE) project is a major project involving integrating unit record data from the Census of Population and Housing with other ABS and non-ABS datasets to create new datasets for statistical and research purposes. The project also adds value to data from the Census of Population and Housing by bringing it together with data from future Censuses.

The CDE project delivers significant public benefits without compromising the privacy of individuals or the confidentiality of their data. The project facilitates:

- improved information to support good government policy making, program evaluation and service delivery; and
- an improved and expanded range of official statistics.

The Australian Statistician announced his intention to proceed with a CDE project in August 2005 after extensive discussion and consultation. The project was first undertaken for the 2006 Census and the ABS intends to continue the project for the 2011 Census. This paper provides an update on the outcomes of the 2006 CDE project and presents plans for the continuation of the project for the 2011 Census.

AUTHORITY

Under the *Australian Bureau of Statistics Act 1975*, the ABS is the central statistical authority for the Australian Government. Among its functions, the ABS is required to:

- collect, compile, analyse and disseminate statistics and related information;
- avoid duplication of collection by official bodies, of information for statistical purposes; and
- achieve maximum possible utilisation, for statistical purposes, of information available to official bodies.

The CDE project is consistent with the legislated function of the ABS to maximise the use, for statistical purposes, of information available to official bodies.

The ABS is obligated to comply with provisions in the *Census and Statistics Act 1905* and the *Privacy Act 1988* to respect the privacy of individuals and to protect the confidentiality of their data. The use and release of data from the CDE project is governed by the provisions outlined under both these Acts.

[Previous Page](#)

[Next Page](#)

Census Data Enhancement Project - 2006

[Contents >> Census Data Enhancement Project - 2006](#)



CENSUS DATA ENHANCEMENT PROJECT - 2006

SUMMARY OF THE 2006 CENSUS DATA ENHANCEMENT PROJECT OUTCOMES

The 2006 Census Data Enhancement (CDE) project encompassed three components:

1. creation of a Statistical Longitudinal Census Dataset (SLCD);
2. bringing together 2006 Census data with ABS and non-ABS datasets using name and address during Census processing to undertake quality studies; and
3. bringing together the 5% SLCD with specified non-ABS datasets for statistical and research purposes.

The 2006 CDE project realised five key benefits:

1. Significant improvements in life tables for Aboriginal and Torres Strait Islander Australians have been achieved.
2. Methodologies for statistical linking, and determining the quality of the linked data produced, have been assessed, resulting in improvements for future CDE projects.
3. The feasibility of continuing with a 5% SLCD has been confirmed.
4. It has been confirmed that it is feasible to automate the matching process used between the Census Post Enumeration Survey and the Census to estimate the number of people who were missed in the Census or who were counted more than once, leading to more efficient and effective processes.
5. It is feasible to bring together data from the Department of Immigration and Citizenship's Settlements Database with the 5% SLCD, and this linked dataset can produce valuable information that no other data source currently provides.

Details of these components and the outcomes achieved in 2006 are provided in Appendix 1.

The success of the studies undertaken as part of the 2006 CDE project paves the way for further studies to improve and enhance a range of both ABS and non-ABS data without compromising the privacy of individuals or the confidentiality of their data.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

Census Data Enhancement Project - 2011

[Contents >> Census Data Enhancement Project - 2011](#)



CENSUS DATA ENHANCEMENT PROJECT - 2011

As in 2006, the 2011 Census Data Enhancement (CDE) project will encompass a number of components, see below links. Of these components, 2 and 5 are new incremental changes in 2011 to the 2006 CDE project.

This section contains the following subsection :

- 1 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, for quality studies
- 2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs
- 3 Wave 2 of a 5% Statistical Longitudinal Census Dataset
- 4 Bringing together the SLCD with other datasets without using name and address for statistical and research purposes
- 5 Bringing together 2011 Census data with other datasets without using name and address after Census processing

[Previous Page](#)

[Next Page](#)

Confidentiality and Privacy

[Contents >> Confidentiality and Privacy](#)



CONFIDENTIALITY AND PRIVACY

OVERVIEW

A fundamental aspect of the Census Data Enhancement (CDE) project is the management of confidentiality and privacy.

All personal information used in the CDE project will be kept confidential. The *Census and Statistics Act 1905* guarantees this protection and legally prevents all ABS staff (including temporary employees) from disclosing information in a manner that is likely to enable the identification of a person or organisation. As a result, potentially identifiable personal information used in the CDE project will not be released to any individual or organisation outside the ABS.

The ABS has an excellent record of managing the personal information provided to it under the *Census and Statistics Act 1905* over 100 years of operations. As a Commonwealth agency that has been set up with the function of gathering data from the community about a range of aspects of Australian life, the ABS undertakes its operations in accordance with the *Privacy Act 1988*.

This section describes the processes that the ABS has in place to manage the personal information used in the CDE project. As confidentiality management is fundamental to the ongoing business of the ABS, relevant current ABS practices and procedures in maintaining the confidentiality of the data collected are also described.

- Legislative Protection;
- Destruction of Census forms and name and address information;
- Access to ABS information; and
- Data security.

This section contains the following subsection :

Legislative Protection

Destruction of census forms and name and address information

Appendix

[Contents >> Appendix](#)

APPENDIX 1 CENSUS DATA ENHANCEMENT PROJECT 2006

The 2006 Census Data Enhancement (CDE) project encompassed three components:

1. creation of a 5% Statistical Longitudinal Census Dataset (SLCD);
2. bringing together 2006 Census data with ABS and non-ABS datasets using name and address during Census processing to undertake quality studies; and
3. bringing together the 5% SLCD with specified non-ABS datasets for statistical and research purposes.

CREATION OF A STATISTICAL LONGITUDINAL CENSUS DATASET (SLCD)

A major component of the CDE project is the creation of a Statistical Longitudinal Census Dataset, or SLCD. The first wave of the SLCD was formed with a 5% random sample of the population from the 2006 Census. The second wave of the SLCD will be created by combining the wave one data in the SLCD 2006 with data from the 2011 Census of Population and Housing.

The 5% SLCD is formed using statistical data linking techniques (see Glossary) rather than matching based on name and address. A paper outlining methods for creating the 5% SLCD, results from similar statistical linking projects, and preliminary results of linking can be found in ABS Research Paper: Exploring Methods for Creating a Longitudinal Census Dataset (ABS cat. no. 1352.0.55.076).

A quality study was undertaken as part of the 2006 CDE project which assessed the likely quality of the 5% SLCD. The results of that study found that the 5% SLCD would produce similar or better quality results to a panel survey. For further information, see Assessing the Likely Quality of the Statistical Longitudinal Census Dataset (ABS cat. no. 1351.0.55.026).

Further information about the second wave of the 5% SLCD and plans for the third wave are outlined in Section 3 of this paper.

BRINGING TOGETHER 2006 CENSUS DATA WITH OTHER DATASETS USING NAME AND ADDRESS DURING CENSUS PROCESSING TO UNDERTAKE QUALITY STUDIES

The second component of the 2006 CDE project was the undertaking of several quality studies which involved bringing together data from the 2006 Census with other ABS and non-ABS datasets. The agreement of the custodians of the non-ABS datasets was required for projects using non-ABS datasets. The aim of these studies was to understand and evaluate the quality of ABS statistical operations and outputs, to better inform ABS on the most suitable statistical techniques for bringing together the 5% SLCD and other datasets, and to assess the quality of datasets created using these techniques.

The quality studies were undertaken during the 2006 Census processing period using

names and addresses to undertake the linkage, after which all Census forms and names and addresses held by the ABS were destroyed. The linked datasets created for the quality studies did not contain name and address information and were destroyed at the completion of the studies (see previous comments regarding security, confidentiality and accessibility).

A number of quality studies were proposed for Census 2006 as outlined in the information paper Census Data Enhancement Project: An Update (ABS cat. no. 2062.0). The outcomes of studies that did proceed are outlined below.

Feasibility of combining the 5% SLCD with data from future Censuses

This study aimed to test the feasibility of bringing together a 5% sample of one Census with subsequent Censuses using statistical techniques. This was simulated by linking the 2005 Census Dress Rehearsal dataset to the 2006 Census data both with and without names and addresses as matching variables. The linking using name and address acted as a benchmark for assessing the quality of the linking without using name and address. Details about the linking methodologies used, the application of these methodologies in the quality study using the 2005 Census Dress Rehearsal dataset and the outcomes of the study were released in three research papers:

- The first paper, Research Paper: Methodology of Evaluating the Quality of Probabilistic Linking (ABS cat. no. 1351.0.55.018) was released in April 2007. The paper included a brief description of developments in data linking at the ABS, outlined the data linking methodology and quality measures considered as part of the quality study, and summarised preliminary results based on the Census Dress Rehearsal data.
- The second paper, A Linkage Method for the Formation of the Statistical Longitudinal Census Dataset (ABS cat. no. 1351.0.55.025) was released in August 2009 and described the methods and processes used in the simulated SLCD quality study.
- A complementary research paper Assessing the Likely Quality of the Statistical Longitudinal Census Dataset (ABS cat. no. 1351.0.55.026) described a variety of methods used to examine the quality of the data linked without name and address. The paper included predictions of the quality that can be expected when the first two waves of the 5% SLCD are linked.

The ABS will be undertaking a series of quality studies using the 2010 Census Dress Rehearsal dataset. For further details, see '1 Bringing together 2011 census data with other datasets during census processing using name and address, for quality studies'.

[Back to top](#)

Indigenous Mortality

This study brought together data from the 2006 Census with data from death registrations for the period August 2006 to 30 June 2007, during the Census processing period using name and address. The study assessed the undercoverage of Indigenous deaths in death registration records; identified factors that may be contributing to undercoverage in Indigenous deaths in death registrations; and assessed the feasibility of calculating and applying adjustment factors to improve estimates of Indigenous mortality.

Findings from the quality study using death registration data can be found in the information paper, Census Data Enhancement - Indigenous Mortality Quality Study, 2006-07 (ABS cat. no. 4723.0).

Adjustment factors obtained in the quality study were subsequently used in the development

and introduction of a new method to derive adjusted Indigenous deaths used to produce life tables and life expectancy estimates for Aboriginal and Torres Strait Islander Australians. The availability of information from the quality study considerably improved the quality and robustness of the estimates of Indigenous life expectancy. These estimates had previously relied on a range of assumptions about the level of under identification of Indigenous deaths in each jurisdiction. The linked Census and death registration data pointed to significant deficiencies in these assumptions, resulting in significant underestimation of Indigenous life expectancy in some states/territories. For further information, see Discussion Paper: Assessment of Methods for Developing Life Tables for Aboriginal and Torres Strait Islander Australians, 2006 (ABS cat. no. 3302.0.55.002).

The ABS will be undertaking an Indigenous Mortality project as part of the 2011 CDE project. For further details, see '2.1 Indigenous Mortality Project'.

Assessing the feasibility of bringing together data from the Department of Immigration and Citizenship's Settlement Database and the 2006 Census

The Migrants Quality Study was conducted to assess the feasibility of linking the Department of Immigration and Citizenship's Settlement Database (SDB) to the 5% Statistical Longitudinal Census Dataset (SLCD) without the use of name and address as linking variables. Findings of the quality study Assessing the Quality of Linking Migrant Settlement Records to Census Data (ABS cat. no. 1351.0.55.027) were released in August 2009. The results from the quality study indicated that linking the SDB to the 5% SLCD is feasible and can produce useful information that no other data source currently provides. However, some quality issues were identified and further work was proposed to ensure that the linked data are correctly interpreted and appropriately used. For further details, see '1 Bringing together 2011 census data with other datasets during census processing using name and address, for quality studies'.

Assessing Automatic Data Linking for the Census Post Enumeration Survey

The Census Post Enumeration Survey (PES) is conducted a few weeks after the Census to estimate the number of people who were missed in the Census or who were counted more than once. This is done by matching Census and PES responses. A quality study undertaken after the 2006 PES assessed the feasibility of introducing automated linking processes to improve the efficiency and effectiveness of the PES, in line with previous changes to the survey, such as the introduction of Computer Assisted Interviewing in 2006. The study found that automated data linking can provide important quality and efficiency gains to replace in part, though not entirely, the clerical matching process. A key advantage of automated data linking is that it provides a greater capability for locating respondents at undisclosed or poorly reported Census night addresses.

[Back to top](#)

BRINGING TOGETHER THE 5% SLCD WITH SPECIFIED NON-ABS DATASETS FOR STATISTICAL AND RESEARCH PURPOSES

The third component of the 2006 CDE project was the bringing together of the 5% SLCD with specified non-ABS datasets using statistical techniques. The agreement of the custodians of the non-ABS datasets was required for these projects, and projects could only be for statistical and research purposes.

Three non-ABS datasets were identified for this component of the CDE Project. These were birth and death register data, including cause of death data; migrants data from the Department of Immigration and Citizenship's Settlement Database (SDB); and national disease registers.

A statistical study based on bringing together data from the Department of Immigration and Citizenship's Settlement Database with data from the 2006 Census was undertaken. The bringing together of migrant information with Census information had the potential to provide insights into patterns of settlement of different groups of migrants, including family formation, housing, labour force characteristics, changing occupations, educational pathways and region of settlement. The study started after the completion of the quality study (discussed above) and results have been released Perspectives on Migrants, June 2010 (ABS cat. no. 3416.0).

PRIVACY AND CONFIDENTIALITY

A fundamental aspect of the CDE project is the management of privacy and confidentiality. The ABS applied strict protocols to the 2006 CDE project to ensure that the privacy of individuals and the confidentiality of their data was protected throughout the 2006 CDE project. These protocols included:

- legislative protections requiring all data collected by, or supplied to the ABS, including datasets created for the CDE project, to remain confidential to the ABS;
- destruction of all Census forms and deletion of all name and address information from the 2006 Census, including the names and addresses for the 5% sample of the population contained in the SLCD. The ABS will not retain Census name and address once Census processing is completed. The only exception is if a person explicitly agrees by answering the relevant question on the Census form to have their name-identified responses retained by the National Archives of Australia for release in 99 years time (see Glossary for further detail);
- strict application of standard ABS procedures to ensure that all aggregate outputs disseminated by the ABS as a result of the quality studies were unlikely to enable identification of any individual or household.
- meeting all obligations under the Information Privacy Principles for the 2006 Census, including the provision of information about the purpose and use of the information provided in the Census, data security and the release of a CDE Fact Sheet.

Audits undertaken by Oakton in 2008 and 2010 found that:

- the security and confidentiality arrangements put in place to protect the data associated with the Statistical Longitudinal Census Dataset (SLCD) and the quality studies complied with the originally proposed controls; and
- the linked datasets created using name and address were deleted after use.

[Back to top](#)

[Previous Page](#)

1 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, for quality studies

[Contents >> Census Data Enhancement Project - 2011 >> 1 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, for quality studies](#)

1 BRINGING TOGETHER 2011 CENSUS DATA WITH A SMALL NUMBER OF PREDETERMINED DATASETS DURING CENSUS PROCESSING USING NAME AND ADDRESS, FOR QUALITY STUDIES

The ABS is intending to undertake a small number of predetermined data integration projects associated with the 2011 Census of Population and Housing which will involve linking the Census to ABS and non-ABS datasets using name and address, for quality studies.

Each of these studies has been vetted to ensure the studies demonstrate significant potential community benefit; are statistically appropriate; cannot be done without name and/or address to bring the data together; and do not require identifiable information to be accessible outside the ABS. After careful consideration, the Australian Statistician has determined that each of these studies should proceed.

The key features of these studies are:

- once the purpose of each study has been met, all linked datasets will be destroyed;
- the studies can only be undertaken during the Census processing period when name and address are available;
- the linked datasets created through these projects will not leave the ABS and will only be accessible by those ABS officers directly involved in the study; and
- the ABS will not retain Census names and addresses once Census processing is completed. The only exception is if a person explicitly agrees by answering the relevant question on the Census form to have their name-identified responses retained by the National Archives of Australia for release in 99 years time (see Glossary for further detail);

What has changed since the 2006 Census

The 2006 CDE project made use of name and address information from the 2006 Census during the Census processing period to undertake a number of identified quality studies (see Appendix 1). These quality studies have demonstrated the potential for significant statistical improvements to be made by bringing together data from the Census with other datasets. As noted in Appendix 1, this included immigration and mortality data. For the 2011 Census, the datasets proposed to be temporarily brought together with the Census for quality studies are listed below.

There will be no change to the methodologies used in 2006 to bring together datasets in 2011. Names and addresses collected on the 2011 Census will be destroyed at the end of Census processing.

The purpose of each quality study to be undertaken during the 2011 Census processing period are to:

- assess the quality of and provide a benchmark for, data integration projects that do not make use of name and address; and
- assess the data integration methodology to help in improving processes for future data integration projects.

Four quality studies will be undertaken for the 2011 Census. These are:

1. Census 2011 Dress Rehearsal to 2011 Census data to provide a benchmark standard

for the 5% SLCD;

2. 2011 Census to Department of Immigration and Citizenship's Settlements Database to compare linkage outcomes and quality with earlier work linking the 2006 Census to the Settlements Database, enabling evaluation of subsequent improvements implemented by DIAC to the Settlements database and to evaluate opportunities for further development. The study will also provide a benchmark for projects which bring together the Settlements Database with both the 5% SLCD and the 2011 Census data without using name and address;
3. studies related to an overall strategy to develop an Australian Longitudinal Learning Database: a) 2011 Census to school and early childhood education student enrolment data; b) 2011 Census to a sample of student data from the National Assessment Program; and c) 2011 Census to the Australian Early Development Index; and
4. 2011 Census to a Western Australian Enhanced Mortality dataset to provide input into development of national best practice guidelines for data linkage related to Indigenous people.

Benefits of the Quality Studies

Undertaking these studies provides increased quality assurance for proposed as well as future data integration projects in terms of both the methodologies applied and the quality of the output from the projects.

Data involved in the Quality Studies

These quality studies will temporarily bring together the 2011 Census data with a range of ABS and non-ABS datasets using name and address during Census processing. Names and addresses collected on the Census will be destroyed at the end of Census processing and the linked datasets created for the quality studies will be deleted once the purpose of the project has been fulfilled.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs

[Contents >> Census Data Enhancement Project - 2011 >> 2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs](#)

2 BRINGING TOGETHER 2011 CENSUS DATA WITH A SMALL NUMBER OF PREDETERMINED DATASETS DURING CENSUS PROCESSING USING NAME AND ADDRESS, TO CREATE STATISTICAL OUTPUTS

The ABS is intending to undertake two data integration projects associated with the 2011 Census of Population and Housing which will involve linking the Census to ABS and non-ABS datasets using name and address, to create statistical outputs. These are:

2.1 Indigenous Mortality Project

2.2 Enhancing Australia's Cancer Statistics Project

Each of these projects has been vetted to ensure the projects demonstrate significant potential community benefit; are statistically appropriate; cannot be done without name and/or address to bring the data together; and do not require identifiable information to be accessible outside the ABS. After careful consideration, the Australian Statistician has determined that each of these studies should proceed.

The key features of these projects are:

- once the purpose of each project has been met, all linked datasets will be destroyed;
- the projects can only be undertaken during the Census processing period when name and address are available;
- the linked datasets will not leave the ABS and will only be accessible by those ABS officers directly involved in the project; and
- the ABS will not retain Census name and address once Census processing is completed. The only exception is if a person explicitly agrees by answering the relevant question on the Census form to have their name-identified responses retained by the National Archives of Australia for release in 99 years time (see Glossary for further detail).

What has changed since the 2006 Census

The 2006 Census Data Enhancement (CDE) project made use of name and address information from the 2006 Census during the Census processing period to undertake a number of identified quality studies (see Appendix 1). These quality studies have demonstrated the potential for significant statistical improvements to be made by bringing together data from the Census with other datasets. The outputs of the quality studies in 2006 were used to improve statistics and test and report on linking methodology. After careful consideration, the Australian Statistician has decided to harness these improvements to create statistical outputs in aggregate form.

There will be no changes to the methodologies used in 2006 to bring together the datasets. Once the projects have fulfilled their purpose the linked datasets will be destroyed. Names and addresses collected on the Census will be destroyed at the end of Census processing.

The purpose of the two proposed data integration projects to be undertaken during the 2011 Census processing period to create statistical outputs is outlined in the links provided above.

[Back to top](#)

This section contains the following subsection :

[2.1 Indigenous Mortality Project](#)

[2.2 Enhancing Australia's Cancer Statistics Project](#)

[Previous Page](#)

[Next Page](#)

3 Wave 2 of a 5% Statistical Longitudinal Census Dataset

[Contents >> Census Data Enhancement Project - 2011 >> 3 Wave 2 of a 5% Statistical Longitudinal Census Dataset](#)

3 WAVE 2 OF A 5% STATISTICAL LONGITUDINAL CENSUS DATASET

An important feature of the Census Data Enhancement (CDE) project is the formation of a Statistical Longitudinal Census Dataset (SLCD) by bringing together data from the 2006 Census with data from the 2011 Census and future Censuses to build a picture of how society moves through various changes: which groups are affected by different types of change and in what way.

Wave 1 of the SLCD was created from the 2006 Census dataset by selecting a random sample of 5% of persons in the 2006 Census of Population and Housing. Wave 2 of the SLCD will endeavour to bring together the wave 1 records with their corresponding records in the 2011 Census.

Subsequent waves will be created with each new Census, providing a longitudinal dataset of information about 5% of the Australian population.

At each Census, the 5% SLCD will be augmented with a 5% sample of children who have been born and immigrants who have arrived since the previous Census. There will also be some provision for topping up the sample to maintain a dataset that is consistently 5% of the Census population at any point in time.

The third wave of the 5% SLCD will be created in 2016. For this wave, the ABS will make use of a non-identifying grouped numeric code based on name to improve the accuracy of the linked dataset as well as improve the efficiency of the linking process. The decision to use a non-identifying grouped numeric code is based on the outcomes of a 2006 CDE quality study which investigated the statistical techniques used to undertake data linkage and evaluated the feasibility of creating the 5% SLCD without using name and address. The study demonstrated that, in the absence of name and address, inclusion of a non-identifying grouped numeric code when linking records can improve accuracy and efficiency. For further information, see *Assessing the Likely Quality of the Statistical Longitudinal Census Dataset* (ABS cat. no. 1351.0.55.026).

The non-identifying grouped numeric code will be assigned to all records in the 5% SLCD dataset from 2011. It will be created from a combination of letters from first and last names using a secure one-way process, meaning that it cannot be reversed to identify individuals. Each code will represent approximately 2000 people and therefore will not be unique to an individual. The code will only be accessible to those ABS staff creating the linked dataset, and will not be released outside the ABS.

The non-identifying grouped numeric code will be used in conjunction with characteristics such as age, sex, geographic region and country of birth to link records from the 5% SLCD to the 2016 Census and future Censuses using probabilistic record linkage techniques. Name and address information will not be used in the linkage process and will not be available for the 5% SLCD dataset as they are deleted at the end of Census processing.

What has changed since the 2006 Census

The formation of an SLCD was foreshadowed in 2005. The addition of a second wave of Census data to the 5% SLCD from 2006 will provide the first longitudinal view of the Census, for statistical and research purposes. The retention of a non-identifying grouped numeric code on the 5% SLCD to assist the data linking process for the future is a change being made to improve accuracy and efficiency. The code is not an identifier and does not add privacy risk.

Benefits of the Statistical Longitudinal Census Dataset

Each five-yearly Census provides a rich set of information about Australian people and households at a point in time. It provides information on topics such as family structure; education and qualifications; presence of a severe or profound disability; work, including hours worked, occupation and industry; income and housing; country of birth; year of arrival and indigenous status. It is able to provide a rich picture of social and economic conditions at a particular point in time, and how these conditions are changing over time and across population groups.

What the Statistical Longitudinal Census Dataset (SLCD) adds to this, is the ability to study patterns in how social and economic conditions change over time at the individual level, and provide insight into the pathways that tend to lead to particular outcomes, and how these pathways vary for different population groups. It also enables the study of likely consequences of certain socio-economic circumstances for different population groups, in terms of the likely outcomes as evidenced by the patterns in the longitudinal data. It can help develop strategies to achieve positive pathways, and avoid negative ones, and can help policy makers in assessing both the social and financial benefits of related intervention strategies.

As well as using the longitudinal Census data in its own right, the very large Census sample can be used to help inform on the quality of transition probabilities measured in more frequent smaller longitudinal studies, and, particularly for sub population groups may allow adjustment mechanisms to improve the socio-economic modelling that frequently underpins government policy making and research.

The 5% SLCD containing 2006 and 2011 Census data will be available for statistical analysis and research purposes from 2013. Standard ABS confidentiality methods will be applied and the data will be accessible through standard ABS secure data access arrangements. No information that is likely to enable identification of an individual or household will be released (See 'Confidentiality and Privacy').

Data involved in the Statistical Longitudinal Census Dataset

The creation of the 5% SLCD itself only involves the use of data from the Census of Population and Housing.

The 2006 SLCD dataset and the 2011 Census dataset will be brought together using a statistical method referred to as 'probabilistic record linkage'. This involves bringing together data from the two datasets without using names and addresses but by using a number of characteristics common to both datasets such as age, sex, geographic region and country of birth. All possible linkages based on these data items are evaluated and the records for which the linkage is most likely to be correct are brought together. For many individuals this linkage would be correct while for some others it will not. Some inaccuracy in the linkage will not generally affect statistical conclusions drawn from the linked data, although care does need to be taken in the interpretation of results.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

4 Bringing together the SLCD with other datasets without using name and address for statistical and research

purposes

Contents >> Census Data Enhancement Project - 2011 >> 4 Bringing together the SLCD with other datasets without using name and address for statistical and research purposes

4 BRINGING TOGETHER THE SLCD WITH OTHER DATASETS WITHOUT USING NAME AND ADDRESS FOR STATISTICAL AND RESEARCH PURPOSES

The 5% Statistical Longitudinal Census Dataset (SLCD) can be enhanced further by bringing it together with specified non-ABS datasets using statistical techniques.

What has changed since the 2006 Census

The 2006 Census Data Enhancement (CDE) project enabled the bringing together of the 5% SLCD with specified non-ABS datasets. There is no change to this intention. However, as for the linkage associated with the 5% SLCD dataset, it is intended to use a non-identifying grouped numeric code together with such data items as sex, date of birth and country of birth to link the 5% SLCD and non-ABS datasets. Name and address will not be used in bringing the datasets together.

It is not intended to bring together the 5% SLCD with ABS household survey data as the overlapping sample with the 5% SLCD would be too small to be useful.

Benefits of linking the SLCD with other non-ABS datasets

Some important data able to be used for statistical and research purposes are not collected by the ABS. However, in performing its functions under the *Australian Bureau of Statistics Act 1975*, once the data is supplied to the ABS the data is legally protected by the *Census and Statistics Act 1905* which requires the ABS to keep information provided to it confidential.

Integrating the 5% SLCD with other non-ABS datasets can significantly enhance the statistical value of the SLCD. One example can be found in linking the 5% SLCD with the Department of Immigration and Citizenship's Settlement Database (SDB). From the Census information it is possible to tell whether someone was born in Australia or overseas, and in the latter case, the year of arrival in Australia. However, linking the SDB to the 5% SLCD, adds information on the type of visa used. As the socio-economic circumstances, areas of policy concern, and related policy implications for different visa types are very different, the ability to analyse the groups separately is important.

A quality study undertaken as part of the 2006 Census Data Enhancement (CDE) project found that linking the SDB to the 5% SLCD is feasible and will indeed produce useful information that no other data source currently provides (see Research Paper: Assessing the Quality of Linking Migrant Settlement Records to Census Data (ABS cat. no. 1351.0.55.027)).

Data involved in linking the SLCD with specified non-ABS datasets

The 2006 CDE project supported the bringing together of the 5% SLCD with three non-ABS datasets:

- Birth and death register data, including cause of death data;
- Settlement Database data; and
- National Disease Registers (this project did not proceed).

A statistical study linking the 2006 5% SLCD and the Settlement Database was undertaken

with results released in June 2010. See Appendix 1 for more details.

At this early stage of the 2011 CDE project, no data integration projects involving the 5% SLCD are planned.

Any linked datasets created by bringing the 5% SLCD data together with a non-ABS source will be retained within the ABS. Standard ABS confidentiality methods will apply and the linked datasets will be accessible through standard ABS secure data access arrangements. These arrangements ensure that no information likely to enable the identification of an individual will be released.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

5 Bringing together 2011 Census data with other datasets without using name and address after Census processing

[Contents](#) >> [Census Data Enhancement Project - 2011](#) >> **5 Bringing together 2011 Census data with other datasets without using name and address after Census processing**

5 BRINGING TOGETHER 2011 CENSUS DATA WITH OTHER DATASETS WITHOUT USING NAME AND ADDRESS AFTER CENSUS PROCESSING

In 2011, the ABS will support certain projects that bring together the full 2011 Census dataset with other ABS and non-ABS datasets without name and address after Census processing. Projects of this sort were not conducted with the 2006 Census, where the non-ABS datasets were brought together with the 5% Statistical Longitudinal Census Dataset file.

Any data integration projects linking datasets to the full Census without name and address will need to satisfy the following:

- name and address will not be used in the linking;
- the link will be to a single cycle of the Census only;
- the use of the Census data must be for statistical and research purposes only;
- the use of Census data must be of significant community benefit;
- the use of Census data is appropriate to address the research questions under consideration; and
- the use of Census data must not require access to identifiable information.

All such data integration projects will require the approval of the Australian Statistician to proceed.

Details of all such projects approved to proceed will be published. Datasets created for these projects will have standard ABS confidentiality methods applied and be accessible through standard ABS secure data arrangements whereby no information likely to enable identification of an individual will be released.

Legislative Protection

[Contents](#) >> [Confidentiality and Privacy](#) >> [Legislative Protection](#)

Legislative Protection

All ABS officers are bound by strict secrecy provisions under the *Census and Statistics Act 1905*. Officers sign an undertaking of fidelity and secrecy to ensure that they are aware of their responsibilities. Section 19 of the *Census and Statistics Act 1905* forbids past or present ABS officers from divulging information collected under this Act, either directly or indirectly, under penalty of up to 120 penalty units (currently \$13,200) or imprisonment for 2 years or both.

In this manner, the *Census and Statistics Act 1905* protects the confidentiality of data provided to the ABS. These protections apply to all data collected by, or supplied to the ABS, including the data to be used for the Census Data Enhancement (CDE) project.

In particular, the ABS will ensure the full protection of the *Census and Statistics Act 1905* is applied to any dataset created through this project. This Act requires the ABS to keep information provided to it confidential. Potentially identifiable data created through the CDE project will not be provided outside of the ABS.

Destruction of census forms and name and address information

[Contents](#) >> [Confidentiality and Privacy](#) >> [Destruction of census forms and name and address information](#)

Destruction of census forms and name and address information

ABS destroys Census forms after statistical processing has been completed. Paper is pulped and recycled.

The eCensus is the electronic option for returning Census forms, which allows completion of the Census via the Internet. To ensure that the information is delivered safely to the ABS, the strongest encryption technology that current browsers will support is used. eCensus data sent to the ABS via the Internet is not able to be read by anyone other than the ABS. At the end of the Census, the hard disk drives used to store information will be wiped under the supervision of the ABS to ensure there is no possibility of any Census data being accessed by any unauthorised person.

The ABS will not retain Census name and address once Census processing is completed. The only exception is if a person explicitly requests that their data is to be archived, by answering the relevant question on the Census form to have their name-identified responses retained by the National Archives of Australia for release in 99 years time (see Glossary for further detail). The ABS does not retain copies of this information.

Access to ABS information

[Contents](#) >> [Confidentiality and Privacy](#) >> [Access to ABS information](#)

Access to ABS information

Procedures are put in place to ensure all aggregate outputs disseminated by the ABS are not likely to enable the identification of a particular person.

The *Census and Statistics Act 1905* allows the ABS to disclose confidentialised unit record data (that is data that is not likely to enable the identification of a particular person) only for statistical purposes, at the discretion of the Australian Statistician. When this occurs, the records are disclosed in the form of Confidentialised Unit Record Files (CURFs). These files have had all identifying information removed and, in addition, the data items that may be likely to enable the identification of individuals are only released in broad categories (for example location may only be released at the State/Territory level). Furthermore, more advanced confidentialisation occurs through checking the CURFs for records with uncommon combinations of responses. These records may be altered slightly to ensure individual responses cannot be identified.

All users who request access to these confidentialised datasets must state what their intended statistical purpose is for using the data, and sign an undertaking to keep the data secure and not to attempt to identify an individual. Should a user breach the conditions of the undertaking and deliberately attempt to identify an individual, they are subject to prosecution under the *Census and Statistics Act 1905*.

Access to confidentialised unit record data from the 5% SLCD and datasets created by bringing together, without name and address, the Census dataset or 5% SLCD with other datasets, would be subject to all the above procedures. Data will be available through standard ABS secure data access arrangements.

Access to temporary datasets created by bringing together, with name and address, the Census dataset with other datasets will be restricted to a small number of ABS officers on a needs to know basis. These officers are bound by strict secrecy provisions under the *Census and Statistics Act 1905*. Officers sign an undertaking of fidelity and secrecy to ensure that they are aware of their responsibilities. Section 19 of the *Census and Statistics Act 1905* forbids past or present ABS officers from divulging information collected under this Act, either directly or indirectly, under penalty of up to 120 penalty units (currently \$13,200) or imprisonment for 2 years or both.

Personal privacy is paramount at the ABS. The Australian community can be confident that the ABS will keep their personal information secure - including data provided on paper Census forms or in the eCensus. The ABS has never and will never release identifiable personal information to any outside organisation, agency or project.

By the law outlined above, organisations such as the Tax Office and credit reference groups cannot have access to personal details from the Census or Census Data Enhancement projects.

Data security

[Contents](#) >> [Confidentiality and Privacy](#) >> [Data security](#)

Data security

The ABS maintains practices of a high standard to ensure the security of all information it holds. Features of the ABS environment are:

- strong security arrangements for all ABS information technology systems. ABS conforms with IT Security arrangements set out in the Australian Government Information Security Manual ASCI 33;
- strict control of access to all ABS premises in accordance with the Commonwealth Protective Security Manual to ensure compliance with legislative responsibilities;
- appropriate personnel security arrangements. Upon appointment all ABS staff undergo security checks and are required to sign an undertaking of fidelity and secrecy;
- a secured Internet gateway which is reviewed annually by Defence Signals Directorate;
- regular Protective Security risk reviews to ensure that security arrangements continue to be effective; and
- an ongoing program of security audits and reviews of computer systems and the physical environment.

In addition the ABS induction and training strategy for its staff places strong emphasis on the importance of security in safeguarding confidentiality, and on the appropriate use of the technology environment.

2.1 Indigenous Mortality Project

[Contents](#) >> [Census Data Enhancement Project - 2011](#) >> [2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs](#) >> [2.1 Indigenous Mortality Project](#)

2.1 Indigenous Mortality Project

As part of the Australian Government's Closing the Gap strategy, the ABS will deliver aggregated statistical information to improve estimates of Indigenous life expectancy.

Death registrations data are a key input for estimating life expectancy and are provided to the ABS as an administrative dataset by the State and Territory Registrars of Births, Deaths and Marriages. Whilst most Indigenous deaths are registered, for some the Indigenous status is not correctly identified, or identified inconsistently with respect to the Indigenous status in the Census.

A Census Data Enhancement (CDE) quality study undertaken in 2006 showed that estimates of Indigenous life expectancy could be significantly improved by adjusting for differences between the Indigenous status in death registrations and the Census and Post Enumeration Survey. As such, adjustment factors obtained in the 2006 quality study were used to derive adjusted Indigenous deaths for use in compiling life tables and life expectancy estimates for Aboriginal and Torres Strait Islander Australians. For further information see Experimental Life Tables for Aboriginal and Torres Strait Islander Australians, 2005-2007 (ABS cat. no. 3302.0.55.003). This same method will be applied for the 2011 Census.

The 2011 Indigenous Mortality Project will:

- assess the consistency of Indigenous status as reported in death registration and Census data;
- estimate measures of undercoverage of Indigenous deaths by state/territory and remoteness areas of Australia;
- investigate the feasibility of applying adjustment factors for Indigenous deaths output data; and
- provide input into the compilation of Indigenous life tables, life expectancy estimates and Indigenous/non-Indigenous differences in life expectancy and other mortality measures, that are consistent with population estimates based on the adjusted 2011 Census of Population and Housing.

Benefits of the Indigenous Mortality Project

The benefits of the Indigenous Mortality project in 2006 are outlined in Appendix 1. In addition to the benefits achieved in 2006, repeating this project will enable comparable life expectancy estimates between 2006 and 2011 to be produced by using a consistent methodology.

The project enables reporting against the COAG target "to close the life expectancy gap within a generation".

The project also provides information on the quality of COAG performance indicators relating to the mortality rates of Indigenous and non-Indigenous people.

More broadly, the project provides information to inform strategies for improving Indigenous identification in administrative data.

Data involved in the Indigenous Mortality Project

This project will temporarily link the 2011 Census data to death registration data using names and addresses during the Census processing period. All death registrations from August 2011 to August 2012 will be linked to the 2011 Census. Once the purpose of the project has been fulfilled, all linked datasets will be destroyed.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

2.2 Enhancing Australia's Cancer Statistics Project

Contents >> Census Data Enhancement Project - 2011 >> 2 Bringing together 2011 Census data with a small number of predetermined datasets during Census processing using name and address, to create statistical outputs >> 2.2 Enhancing Australia's Cancer Statistics Project

2.2 Enhancing Australian Cancer Statistics Project

The ABS is looking to undertake a project temporarily bringing together 2011 Census data with data from the Australian Cancer database maintained by the Australian Institute of Health and Welfare to:

1. to develop additional national cancer statistics not currently available; and
2. to gain a quality measure of the identification of Indigenous status in the national cancer data base and potentially an adjustment factor to improve the estimates of cancer incidence and mortality for the Indigenous community.

Benefits of the Enhance Australian Cancer Statistics Project

There are substantial benefits in improving our understanding of the relationship between socio-economic variables such as income, education, location and housing, on cancer risks and their outcomes. Are some groups more prone to certain cancers, and are some groups showing different levels of mortality resulting from cancer?

The improvement in the identification of Indigenous status will improve knowledge of cancer risks in the Indigenous community. Currently Indigenous status is only provided to Cancer registries via hospitalisation and mortality data, both of which are known to under-enumerate Indigenous persons. By undertaking a cross check with the Census, this will confirm existing Indigenous persons on the cancer data base, identify not previously identified Indigenous persons and also identify Indigenous persons identified through the cancer data base that were not identified in the Census. Adjusted statistics for the Indigenous community may be able to be calculated. This work is consistent with the COAG National Partnership Agreement on "Closing the Gap in Indigenous Health Outcomes".

Data involved in the Enhancing Australian Cancer Statistics Project

This project will temporarily bring together, in the ABS, data from the 2011 Census and the Australian Cancer database maintained by the Australian Institute of Health and Welfare using name and address during the Census processing period. Once the purpose of the project has been fulfilled, all linked datasets will be destroyed.

[Back to top](#)

[Previous Page](#)

[Next Page](#)

Explanatory Notes

Explanatory Notes

EXPLANATORY NOTES

**ABBREVIATIONS
GLOSSARY**

ABBREVIATIONS

ABS Australian Bureau of Statistics
ABSDL Australian Bureau of Statistics Data Laboratory
ALLD Australian Longitudinal Learning Database
CDE Census Data Enhancement
COAG Council of Australian Governments
CURF Confidentialised Unit Record Files
DIAC Department of Immigration and Citizenship
NAA National Archives of Australia
PES (Census) Post Enumeration Survey
SLCD Statistical Longitudinal Census Dataset
WA Western Australia

GLOSSARY

ABS Data laboratory

The ABS Data Laboratory (ABSDL) is the data analysis solution for high-end data users who want to extract full value from ABS microdata. The ABSDL provides an interactive environment, enabling the analysis of Basic, Expanded or Specialist (customised) Confidentialised Unit Record Files (CURFs).

Australian Longitudinal Learning Database

The Australian Longitudinal Learning Database (ALLD) has been proposed as a national longitudinal statistical database on the education pathways and outcomes of Australian students from early childhood education through to the end of their schooling or education. The ALLD would be constructed from administrative records, and, with community support, would include data drawn from Census and survey records.

Birth register data

The responsibility for registration of births in Australia lies with the individual State and Territory Registrars of Births, Deaths and Marriages. A Birth Registration Statement is completed by at least one of the parents of a baby. This information is the basis of the data provided to the ABS for processing and production of birth statistics.

[Back to top](#)

Census processing period

The period of time immediately after the conduct of the Census of Population and Housing during which the Census forms are processed to produce statistical outputs.

Census Time Capsule

In Australian Censuses prior to 2001, forms and other name-identified records have been destroyed once the statistical data required for the purposes of the Census have been extracted.

Following recommendations from the House of Representatives Standing Committee, the Government decided that for the 2001 Census all people would be given the option of having their name-identified responses retained for 99 years (Census Time Capsule). After

99 years, the name-identified data will be made public for future generations. This option was again included in the 2006 Census and will be a permanent feature of future Censuses.

Some 53% of the population chose to have their individual responses from the 2001 Census retained, and 56% from the 2006 Census. These are now with the National Archives of Australia. In order to ensure that the current high levels of public confidence and cooperation in the Census are maintained, and to respect the wishes of those who do not want their information retained for future release, information will only be kept for those persons who explicitly give their consent. For privacy reasons the name-identified information will not be available for any purpose, including by a court or tribunal, within a 99 year closed access period.

After this information has been transferred to the National Archives of Australia and statistical processing is completed, the ABS will destroy all paper and eCensus forms including the computer images of those forms. As in the past, the paper forms will be pulped for recycling.

Confidentialised Unit Record File (CURF)

A CURF is a file of responses to an ABS statistical collection that has had specific identifying information about a person or organisation confidentialised.

The most basic of the techniques employed by the ABS involves ensuring all identifying information, such as names and addresses are not on the files.

Additionally, the data items that are most likely to enable identification of unit records are only released in broad categories. For example, while survey questionnaires may capture your home or business address, microdata may only be released at the State or Territory level.

More advanced confidentialisation occurs through checking the CURFs for records with uncommon combinations of responses. These records may be altered slightly to ensure individual respondents cannot be identified.

Dataset

A file containing the individual responses from a statistical collection, administrative records or register of information (for example disease register). Datasets are used to generate statistical output.

Death register data

Registration of deaths is the responsibility of the individual State and Territory Registrars of Births, Deaths and Marriages and is based on the data provided on an information form. This information form is the basis of the data provided to the ABS for processing and production of death statistics.

Identifiable

In this publication unit record data is considered identifiable if the data available in the record identifies the specific individual to whom it refers.

Longitudinal dataset

A dataset which contains information for the same unit over a number of different points in

time.

Long-term migration data

Statistical data held by the Department of Immigration and Citizenship (DIAC) from the administration of immigration programs. This includes overseas arrivals and departures data, where the period of duration is over 12 months, and visa grant data, including type of visa.

Non-identifying Grouped numeric code

From 2011 a non-identifying grouped numeric code will be included on the 5% SLCD records to improve the accuracy of the linked dataset and the efficiency of the linking process. The code will be based on name and created using a secure one-way process. Each group code will represent about 2000 people.

Statistical Integration Projects

The bringing together of unit record data from different administrative and/or survey sources to provide new datasets for statistical and research purposes. These new datasets address significant research questions, produce new statistical outputs or enable understanding and evaluation of the quality of statistical operations, techniques and/or outputs.

Statistical purposes

Functions related to the compilation, analysis and dissemination of statistics. Statistical purposes precludes use of a dataset for administrative or client management purposes, where there is an impact on specified individuals.

Statistical techniques

In this publication, statistical techniques refer to the method that would be used to bring together different administrative and/or survey sources. The proposed method is often referred to as probabilistic record linkage, which involves bringing together data from two different datasets using a number of characteristics such as name, address, age/date of birth, sex, geographic region, and country of birth. All possible linkages based on these data items, or a subset of them, are evaluated. The records for which the linkage is most likely to be correct are brought together.

WA Enhanced Mortality dataset

The WA Enhanced Mortality dataset involves linking WA death registrations data with a range of health datasets available in the WA Data Linkage System. Indigenous status will be derived for this linked data set using a number of possible business rules.

[Back to top](#)